REMARKS

Claims 1-42 are pending. Claims 1, 12, 20, 31, and 41 are in independent form.

Rejections under 35 U.S.C. § 103(a)

In sustaining the rejections of claims 1, 12, 20, and 31 under35 U.S.C. § 103, the Advisory Action mailed July 18, 2008 contends that:

> "Wical discloses 'a knowledge base that associates terms of the document with categories of a classification system to develop contextual associations for terminology.' See WICAL, '560, col. 4, lines 39-44. Wherein the knowledge base may be read, accessed, and queried by a hardware implementation such as a computer, one of ordinary skill in the art would have been able to readily discern that said knowledge base would read upon a 'machine-readable network of interrelated concepts.'" *See Advisory Action mailed July 18, 2008*, Continuation Sheet.

Applicant would like to point out that the claims 1 and 20 relate to interacting with a user to add concepts to a machine-readable network of interrelated concepts. Claims 12 and 31 relate to interacting with a user to edit concepts in a machine-readable network of interrelated concepts. Applicant is thus not claiming to have invented machine-readable networks of interrelated concepts. Rather, Applicant's claims are directed to approaches to interacting with a user so that concepts can be added or edited.

There is no description of the addition or editing of concepts in the knowledge base of Wical '560. This is perhaps not surprising given that Wical '560 is primarily concerned with using the described knowledge base in searching. *See, e.g., Wical '560*, col. 4, line 39-44; col. 31, line 6-17.

To ensure that this distinction is recognized by the Examiner, claims 1 and 20 have been amended to recite that a new first concept is created in an existing machine-readable network of interrelated concepts to expand the network of interrelated concepts by adding the new first concept to the existing network of interrelated concepts.

The Advisory action mailed July 18, 2008 also contends that:

> "Additionally, Applicant asserts the argument that WICAL '515 fails to disclose the recited feature of 'concept' because individual words or phrases would not read upon the feature of 'normalized semantic representations.' See Amendment, page 17. The Examiner respectfully disagrees in that under the broadest reasonable interpretation, the use of individual words or phrases to define a category would readily read upon the recited feature of a 'normalized semantic representation.' That is, wherein letters are combined to create discernable terms and phrases, said terms and phrases would sufficiently read upon the requirement of 'normalized semantic representations.'" *See Advisory action mailed July 18, 2008*, Continuation Sheet.

Applicant respectfully disagrees for several reasons. For example, claims 1 and 20 recite that a primary term and a related term representing a first concept are both received from a user and added to a network of interrelated concepts. Claims 12 and 31 both recite that the representation of a first concept on a display includes displaying a collection of one or more terms that express the first concept. It is thus clear that that the recited concepts are expressed/ represented using terms.

A combination of letters "to create discernable terms and phrases" does not "read upon" such an expression/representation of concepts using terms. In this regard, Applicant respectfully submits that letters are not terms. Accordingly, combining letters to form terms does not express/represent concepts that comprise normalized semantic representations, as recited.

To ensure that this distinction is recognized by the Examiner, claims 12 and 31 have been amended to recite that a concept is "expressed by a collection of terms." Combining letters to spell terms and phrases thus clearly does not "read upon" expressing concepts.

Indeed, as discussed numerous times before, Wical '515 itself does not consider the terminology described therein to be concepts that comprise normalized semantic representations. Instead, terminology in Wical '515 are individual words or phrases. *See, e.g., Wical '515*, col. 3, line 14-17. This is apparent in the excerpt of Wical '515 at col. 4, line 49-65 *relied upon by the present rejection*. For the sake of convenience, this cited excerpt is now reproduced.

> "For an example input term, 'short-term', the learning system attempts to select a category that best defines the use of the term in the document set. For this example, the document set is generally about short-term financial instruments, including short-term loans and short-term investments, such as stocks and bonds. To learn the term, the learning system may select the high level category 'business and economics' as the learned category for the input term 'short-term.' Although the document set includes themes about 'business and economics' generally (e.g. short-term financial investments), the input term 'short-term' may be more specifically defined. Thus, if categorized in the 'business and economics' category, the term 'short-term' would be learned at too high of a level (e.g. the associated meaning of the 'business and economics' category is too broad for the term 'short-term' used in the context of the document set)." *See Wical '515*, col. 4, line 49-65 (emphasis added).

Moroever, Applicant and Wical '515 are not alone in considering the terminology to be distinguishable from concepts. In support of this contention, submitted herewith is a copy of an overview of the volume entitled "Ontology Learningfrom Text: Methods, Evaluation and Applications," P. Buitelaar et al. (Eds.) IDS Press (2005) (hereinafter "Buitelaar"). Buitelaar brings together a collection of

selected papers from two workshops on ontology learning. *See, e.g., Buitelaar*, page 3. As shown, e.g., in FIG. 1 of Buitelaar and the written description thereof, Buitelaar considers terminology to be distinguishable from concepts.

Accordingly, applicant *again* respectfully requests that the Examiner set forth with particularity why Applicant, Wical '515, and Buitelaar are incorrect in considering terminology to be distinguishable from concepts.

In the absence of such a showing, or other showing as to how Wical '515, Wical '560, Borgida, and Wical '788 can be combined to arrive at the subject matter of claims 1, 12, 20, and 31, applicant respectfully submits that the obviousness rejections are improper and asks that they be withdrawn. Thus, for the reasons set forth above and in the response filed July 1, 2008, Applicant respectfully requests that the rejections of claims 1, 12, 20, 31, and the claims dependent therefrom be withdrawn.

As best understood, neither the cover page nor the continuation sheet of the Advisory Action mailed July 18, 2008 includes any description of the status of claim 41. If Applicant's Remarks in the response filed July 1, 2008 have been persuasive, Applicant respectfully requests that this be indicated. If Applicant's Remarks were not persuasive, Applicant respectfully requests to be advised as to why.

In the absence of additional comments, Applicant respectfully submits that claim 41 and its dependencies is not obvious over Borgida, Wical '515, and Wical '560 for at least the reasons set forth in the response filed July 1, 2008. Applicant respectfully requests that the rejections of claims 41 and 42 be withdrawn.

It is believed that all of the pending claims have been addressed. However, the absence of a reply to a specific rejection, issue, or comment does not signify agreement with or concession of that rejection, issue, or comment. In addition, because the arguments made above may not be exhaustive, there may be reasons for patentability of any or all pending claims (or other claims) that have not been expressed. Finally, nothing in this paper should be construed as an intent to concede any issue with regard to any claim, except as specifically stated in this paper, and the amendment of any claim does not necessarily signify concession of unpatentability of the claim prior to its amendment.

Please apply the fee for a Request for Continued Examination, a one-month extension of time, and any other charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: September 2, 2008

/John F. Conroy, Reg. #45,485/
John F. Conroy
Reg. No. 45,485

Fish & Richardson P.C.
PTO Customer No. **26192**
12390 El Camino Real
San Diego, California 92130
Telephone: (858) 678-5070
Facsimile: (858) 678-5099

JFC/jhg
14002858.doc

Ontology Learning from Text: Methods, Evaluation and Applications

Edited by

Paul Buitelaar

German Research Center for Artificial Intelligence (DFKI) GmbH, Language Technology Lab & Competence Center Semantic Web, Saarbrücken, Germany

Philipp Cimiano

Institute of Applied Computer Science and Formal Methods (AIFB), Department of Knowledge Management, University of Karlsruhe, Germany

Bernardo Magnini

Center for Scientific and Technological Research (ITC-irst), Cognitive and Communication Technologies Division (TCC), Povo-Trento, Italy

IOS Press

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

ISBN 1-58603-523-1
Library of Congress Control Number: 2005927237

Publisher
IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
The Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

Distributor in the UK and Ireland
IOS Press/Lavis Marketing
73 Lime Walk
Headington
Oxford OX3 7AD
England
fax: +44 1865 750079

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

PRINTED IN THE NETHERLANDS

Ontology Learning from Text: Methods, Evaluation and Applications
P. Buitelaar et al. (Eds.)
IOS Press, 2005

Ontology Learning from Text: An Overview

Paul Buitelaar [a,1], Philipp Cimiano [b] and Bernardo Magnini [c]
[a] *DFKI, Language Technology Lab*
[b] *AIFB, University of Karlsruhe*
[c] *ITC-irst, Centro per la Ricerca Scientifica e Tecnologica*

Keywords. Ontology Learning, Knowledge Acquisition, Text Mining

1. Introduction

This volume brings together a collection of extended versions of selected papers from two workshops on ontology learning, knowledge acquisition and related topics that were organized in the context of the European Conference on Artificial Intelligence (ECAI) 2004 and the International Conference on Knowledge Engineering and Management (EKAW) 2004.

The volume presents current research in ontology learning, addressing three perspectives: *methodologies* that have been proposed to automatically extract information from texts and to give a structured organization to such knowledge, including approaches based on machine learning techniques; *evaluation* methods for ontology learning, aiming at defining procedures and metrics for a quantitative evaluation of the ontology learning task; and finally *application* scenarios that make ontology learning a challenging area in the context of real applications such as bio-informatics.

According to the three perspectives mentioned above, the book is divided into three sections, each including a selection of papers addressing respectively the methods, the applications and the evaluation of ontology learning approaches. However, all selected papers pay considerably attention to the evaluation perspective, as this was a central topic of the ECAI 2004 workshop out of which most of the papers in this volume originate.

2. Ontology Learning

An ontology is an explicit, formal specification of a shared conceptualization of a domain of interest [18], where formal implies that the ontology should be machine-readable and shared that it is accepted by a group or community. Further, it should be restricted to a given domain of interest and therefore model concepts and relations that are relevant to a particular task or application domain.

[1]Correspondence to: Paul Buitelaar, DFKI GmbH, Stuhlsatzenhausweg 3, Saarbrücken 66123, Germany. Tel.: +49 681 302 5325; Fax: +49 681 302 5338; E-mail: paulb@dfki.de.

Ontologies formalize the intensional aspects of a domain, whereas the extensional part is provided by a knowledge base that contains assertions about instances of concepts and relations as defined by the ontology[1]. The process of defining and instantiating a knowledge base is referred to as *knowledge markup* or *ontology population*, whereas (semi-)automatic support in ontology development is usually referred to as *ontology learning*.

Ontology learning is concerned with knowledge acquisition and in the context of this volume more specifically with knowledge acquisition from text. Obviously, much of the work in this area therefore builds on the large body of work in this direction within NLP, AI, and machine learning. As such, the legitimate question arises if the wheel is not being reinvented. Is ontology learning merely a rehash of existing ideas and techniques under a new name? The answer to this should be: no. Although the aims of knowledge acquisition and ontology learning (from text) are certainly overlapping - in essence the acquisition of explicit knowledge implicitly contained in (textual) data - there are, however, also a number of novel and innovative aspects to ontology learning that sets it apart from much of the previous work in knowledge acquisition:

- Ontology learning is inherently multidisciplinary due to its strong connection with the Semantic Web, which has attracted researchers from a very broad variety of disciplines: knowledge representation, logic, philosophy, databases, machine learning, natural language processing, image processing, etc. In consequence, ontology learning has profited from a massive exchange of ideas and techniques that shaped a somewhat different vision of the knowledge acquisition problem.

- Ontology learning, in the Semantic Web context, is primarily concerned with knowledge acquisition from and for Web content and is thus moving away from small and homogeneous data collections to tackle the massive data heterogeneity of the World Wide Web instead.

- Given the machine learning origins of much of the work in ontology learning, the field is rapidly adapting the rigorous evaluation methods that are central to most machine learning work. Therefore, ontology learning will be impacted by efforts to systematically evaluate and compare approaches on well-defined tasks and with well-defined evaluation measures, thus making it a highly challenging field in which only competitive and demonstrable approaches will survive.

In summary, these aspects indeed establish ontology learning as a new and challenging area in its own right, with a lot of innovating research to which also this volume hopes to contribute.

3. The Ontology Learning Layer Cake

A large collection of methods for ontology learning from text have developed over recent years as witnessed by the proceedings of various workshops in this area, e.g. at

[1] Also known in previous work on knowledge representation as T-box and A-box respectively.

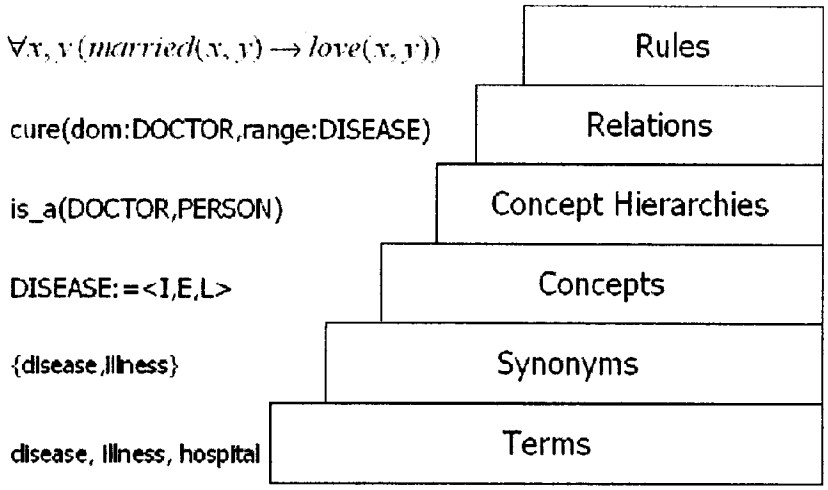

Figure 1. Ontology Learning Layer Cake

ECAI 2000[2], IJCAI 2001[3], ECAI 2002[4], ECAI 2004[5]. Unfortunately, there is not much consensus within the ontology learning community on the exact task they are concerned with, which makes a comparison of approaches difficult[6]. It is therefore the goal of this volume to contribute to a better understanding of the ontology learning task and to help develop metrics and benchmarks to compare research in this field.

In order to estimate the state-of-the-art in ontology learning, we first need to establish the subtasks that together constitute the complex task of ontology development (either manual or with any level of automatic support). Ontology development is primarily concerned with the definition of concepts and relations between them, but connected to this also knowledge about the symbols that are used to refer to them. In our case this implies the acquisition of linguistic knowledge about the terms that are used to refer to a specific concept in text and possible synonyms of these terms. An ontology further consists of a taxonomy backbone (is-a relation) and other, non-hierarchical relations. Finally, in order to derive also facts that are not explicitly encoded by the ontology but could be derived from it, also rules should be defined (and if possible acquired) that allow for such derivations.

All of these aspects of ontology development can be organized in a layer cake of increasingly complex subtasks, as illustrated in Figure 1 (derived from [6]). The example shows the defined knowledge for the concept *disease* and related concepts, i.e. the terms that can be used to refer to or associated with *disease* - also for languages different than English, the taxonomic relation of the concept *doctor* with *person*, a non-hierarchical relation between *doctor* and *disease*, and a rule that can be defined over the *person* and *disease* concepts.

[2]http://ol2000.aifb.uni-karlsruhe.de/

[3]http://ol2001.aifb.uni-karlsruhe.de/home.html

[4]http://www-sop.inria.fr/acacia/WORKSHOPS/ECAI2002-OLT/

[5]http://olp.dfki.de/ecai04/cfp.htm

[6]A start towards surveying the research in this area has been made by the OntoWeb deliverable 1.5 [16]

4. The State-of-the-Art

Given the ontology learning layer cake as discussed above, we can take a closer look at the state-of-the-art in this field. We first examine this layer by layer and then draw some general conclusions at the end.

4.1. Terms

Term extraction is a prerequisite for all aspects of ontology learning from text. Terms are linguistic realizations of domain-specific concepts and are therefore central to further, more complex tasks. The literature provides many examples of term extraction methods that could be used as a first step in ontology learning from text. Most of these are based on information retrieval methods for term indexing [36], but many also take inspiration from terminology and NLP research (see e.g. [2], [14] and [32]).

Term extraction implies more or less advanced levels of linguistic processing, i.e. phrase analysis to identify complex noun phrases that may express terms and dependency structure analysis to identify their internal semantic structure. As such parsers are not always readily available, much of the research on this layer in ontology learning has remained rather restricted. The state-of-the-art is mostly to run a part-of-speech tagger over the domain corpus used for the ontology learning task and then to identify possible terms by manually constructing ad-hoc patterns (e.g. Cimiano et al. and Sabou, both in this volume), whereas more advanced approaches to term extraction for ontology learning build on deeper linguistic analysis as discussed above (e.g. Reinberger and Spyns in this volume and [4]). Additionally, and in order to identify only relevant term candidates, a statistical processing step may be included that compares the distribution of terms between corpora.

4.2. Synonyms and Multilingual Variants

The synonym level addresses the acquisition of semantic term variants in and between languages, where the latter in fact concerns the acquisition of *term translations*. Much of the work in this area has focused on the integration of WordNet[7] for the acquisition of English synonyms, and EuroWordNet[8] for bilingual and multilingual synonyms and term translations. An important aspect of this work is the identification of the appropriate (WordNet/EuroWordNet) sense of the term in question, which determines the set of synonyms that are to be extracted. Obviously, this involves standard word sense disambiguation algorithms, most of which are based on [24] and [42] (see also the SENSEVAL [9] evaluation campaigns for recent approaches on word sense disambiguation). However, specifically in the ontology learning context, researchers have exploited the fact that ambiguous terms have very specific meanings in particular domains allowing for an integrated approach to sense disambiguation and domain specific synonym extraction (compare [5], [22], [30], [31] and [39]).

In contrast to using readily available synonym sets such as provided by WordNet and related lexical resources, researchers have also worked on algorithms for the dy-

[7] WordNet is freely accessible from http://wordnet.princeton.edu

[8] EuroWordNet can be licensed from ELDA at http://www.elda.fr

[9] http://www.senseval.org/

namic acquisition of synonyms by clustering and related techniques. On this basis much work has been done on synonym acquisition from text corpora that is based on Harris' distributional hypothesis that terms are similar in meaning to the extent in which they share syntactic contexts [19], see e.g. [21], [26], [27] and Reinberger and Spyns (this volume). Related work originates out of term indexing for information retrieval, e.g. the family of Latent Semantic Indexing algorithms (LSI, LSA, PLSI and others). LSI and related approaches apply dimension reduction techniques such as those described in [38] or [23] to reveal inherent connections between words, thus leading to group formation. In fact, LSA/LSI-based techniques are especially interesting as they do not run into data sparseness problems such as approaches relying on raw data.

Finally, given the growing importance of the web in knowledge acquisition, there seems to be a current trend to use statistical information measures defined over the web in order to detect synonyms, e.g. [1], [40].

4.3. Concepts

The extraction of concepts from text is controversial as it is not clear what exactly constitutes a concept. In our view, concept induction or formation should provide:

- an intensional definition of the concept
- a set of concept instances, i.e. its extension
- a set of linguistic realizations, i.e. (multilingual) terms for this concept

Thus, we define a concept as a pair with lexicon $(\Im, \Sigma) \oplus L$ where $\Im$ is the intension of the concept, Σ its extension and L describes its linguistic realization. The latter may include complex structures as described in [3].

Most of the research in concept extraction addressed the question from a linguistic or textual perspective, regarding concepts as clusters of related terms. Obviously, this approach overlaps almost completely with that of term and synonym extraction as discussed above.

Alternatively, researchers have addressed the problem from an extensional point of view, e.g. [12] derived hierarchies of named entities from text and thus also discovering concepts from an extensional point of view. The Know-It-All system [11] also aims at learning the extension of concepts such as for example all movie actors appearing on the Web. In the approach of [12] the concepts as well as the extension are derived simultaneously, while [11] essentially populates existing concepts with instances. Note that in this respect, ontology population is very much related to ontology learning.

Finally, intensional concept learning includes the extraction or acquisition of formal and informal definitions. An informal definition might be a textual description, i.e. a gloss of the concept. A formal definition includes the extraction of concept properties, part of which is the extraction of relations between a particular concept and other concepts. The extraction of informal concept definitions is quite rare. In fact the only work reported in this area is the OntoLearn system (Velardi et al. in this volume) that derives WordNet-like glosses for domain-specific concepts. The extraction of formal concept definitions, as far as relation extraction is concerned will be discussed in the next two sections.

4.4. Taxonomy

There are currently three main paradigms exploited to induce taxonomies from textual data. The first one is the application of lexico-syntactic patterns to detect hyponymy relations as proposed by [20]. However, it is well known that these patterns occur rarely in corpora. Thus, though approaches relying on lexico-syntactic patterns have a reasonable precision, their recall is very low. Related to this are also approaches that exploit the internal structure of noun phrases to derive taxonomic relations between classes expressed by the head of the noun phrase and its subclasses that can be derived from a combination of the head and its modifiers [4].

The second paradigm is again based on Harris' distributional hypothesis, as discussed above in the context of synonym extraction and term clustering. In this line, people have mainly exploited hierarchical clustering algorithms to automatically derive term hierarchies from text, e.g. [7], [13], [17].

The third paradigm stems from the information retrieval community and relies on a document-based notion of term subsumption as proposed for example in [37].

4.5. Relations (non-hierarchical)

Recent work on relation extraction from text, other than the is-a relation discussed above, has been addressed primarily within the biomedical field as there are very large text collections readily available (e.g. PubMed[10]) for this area of research. The goal of this work is to discover new relationships between known concepts (i.e. symptoms, drugs, diseases, ...) by analyzing large quantities of biomedical scientific articles (see e.g. [35] [33] [41]).

Most of the work on text mining combines statistical analysis with more or less complex levels of linguistic analysis, e.g. by exploiting syntactic structure and dependencies for relation extraction as reported for instance by [4], [8] and [15]. Relation extraction is therefore also very much related to the problem of acquiring selection restrictions for verb arguments in NLP (compare [34]), as witnessed for instance by the ASIUM system that enables an integrated acquisition of relations between concepts identified in text and so-called sub-categorization frames for the verbs that underlie these relations [13].

Relation extraction through text mining for ontology development was introduced in work on association rules in [29]. Recent efforts in relation extraction from text have been carried on under the ACE (Automatic Content Extraction) program [11], where entities (i.e. individuals) are distinguished from their mentions, and *normalization*, the process of establishing links between mentions in a document and individual entities represented in a ontology, is part of the task for certain kind of mentions (e.g. temporal expressions).

4.6. Rules

The extraction of rules is probably the least addressed researched area in ontology learning. Initial blueprints for this task can be found for example in [25]. Further, the recent PASCAL lexical entailment challenge[12] [10] represents a related problem. In fact, this

[10] http://www.pubmedcentral.nih.gov/

[11] http://www.itl.nist.gov/iad/894.01/tests/ace/

[12] http://www.pascal-network.org/Challenges/RTE/

challenge has strongly increased the awareness of the problem of deriving lexical entailment rules and lead many researchers to address the problem, so that a plethora of approaches to tackle the problem of learning ontological rules from text corpora can be expected in the near future. The main focus hereby has been to learn lexical entailments for application in question answering systems, see [9].

5. The Papers in this Volume

5.1. Methods

The papers by Ryu and Choi, Reinberger and Spyns, Kavalec and Svatek, and Cimiano et al. present methods addressing various aspects of the ontology learning layer cake. Ryu and Choi present an approach to term and taxonomy extraction based on information theory. Reinberger and Spyns present a non-hierarchical term clustering approach over distributed data and thus address concept formation at the lexical level. Kavalec and Svatek present an approach for labeling relations discovered from a corpus in an unsupervised way. Thus, they are targeting the learning of ontological relations, in particular the lexical aspects involved in naming a certain conceptual relation. Cimiano et al. presents a machine-learning based combination of natural language processing, text mining and information retrieval techniques to learn taxonomic relations from various information sources. They thus tackle the problem of deriving sub-/super-concept relations.

5.2. Evaluation

The papers by Faatz and Steinmetz, Velardi et al., and Porzel and Malaka are primarily concerned with evaluation issues. The Faatz and Steinmetz paper presents a proposal for an evaluation methodology of an ontology learning task that consists of enriching the lexical representation with related words on the basis of collocations extracted from a corpus. It thus deals with the lexical aspects of the formation of concepts. The paper of Porzel and Malaka suggests a task-based evaluation framework for ontology learning systems. In particular, they suggest evaluating an automatically learned ontology as a parameter within a relation tagging task. The paper by Velardi et al. describes the results of a thorough evaluation of a specific ontology learning system: OntoLearn. In fact, the system is evaluated at the levels of term, hypernym (i.e. taxonomy) and non-hierarchical relation extraction. Furthermore they address the intensional aspects in concept formation and present an algorithm for generating WordNet-like glosses (natural language definitions) for domain-specific concepts. The paper includes an evaluation of this approach by domain specialists.

5.3. Applications

Several papers are witness that the field is concerned with real tasks. The focus of the paper by Sabou is to automatically learn web service descriptions from software documentation. She is primarily concerned with deriving a taxonomy from these descriptions. Furthermore, she also discusses some extensions of the evaluation measures originally introduced by [28]. The paper by Rinaldi et al. shows how ontology learning methods can be applied to automatically derive relevant terminology for Knowledge Management

applications. While they are mainly concerned with the extraction of relevant terminology, they also discuss the possibility of arranging these terms taxonomically. The paper further has a focus on using the resulting term hierarchies in the context of tasks related to information access. Nedellec and Nazarenko discuss the application of automatically learned ontologies for information extraction. In particular, they suggest a cyclic process in which ontologies are automatically learned or enriched on the basis of a corpus and then used to bootstrap an information extraction system which in turn populates the ontology with newly derived facts.

6. Outlook: Evaluation of Ontology Learning Methods and Tools

Significant progress in the ontology learning field can only follow from a clear and precise definition of this task, and its subtasks as discussed above, and from a general consensus of the scientific community with respect to an evaluation policy of such tasks. For the future, we therefore advocate the organization of a dedicated series of evaluation campaigns on ontology learning and the definition of corresponding metrics and benchmarks. As the result of this we expect an effect similar to that of the TREC[13] and CLEF[14] evaluation campaigns on information retrieval and access, where a common evaluation methodology made it possible to compare the performances of different systems under the same task. The positive consequence has been a concentrated growth of interest and a tremendous boost in research in these areas over the last decade.

Acknowledgements

Paul Buitelaar and Philipp Cimiano are supported by a grant for the SmartWeb project by the German Ministry of Education and Research (01 IMD01 A). Bernardo Magnini has been supported by a grant for the ONTOTEXT project by the Autonomous Province of Trento. Philipp Cimiano has also been supported by the EU FP5 project Dot.Kom (IST-2001-34038).

References

[1] M. Baroni and S. Bisi. Using cooccurrence statistics & the web to discover synonyms in a technical language. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 5, pages 1725–1728, 2004.

[2] D. Borigault, C. Jacquemin, and M.-C. L'Homme, editors. *Recent Advances in Computational Terminology*. John Benjamins Publishing Company, 2001.

[3] Paul Buitelaar. Semantic lexicons: Between terminology and ontology. In K. Simov and A. Kiryakov, editors, *Ontologies and Lexical Knowledge Bases*, pages 16–24. 2000.

[4] Paul Buitelaar, Daniel Olejnik, and Michael Sintek. A protégé plug-in for ontology extraction from text based on linguistic analysis. In *Proceedings of the 1st European Semantic Web Symposium (ESWS)*, 2004.

[13] http://trec.nist.gov/

[14] http://clef.isti.cnr.it/

[5] Paul Buitelaar and Bogdan Sacaleanu. Extending synsets with medical terms. In *Proceedings of the First International WordNet Conference*, 2002.

[6] P. Cimiano. *Ontology Learning and Population: Algorithms, Evaluation and Applications*. PhD thesis, University of Karlsruhe, 2005. forthcoming.

[7] P. Cimiano, A. Hotho, and S. Staab. Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 2005. to appear.

[8] M. Ciramita, A. Gangemi, E. Ratsch, J. Saric, and I. Rojas. Unsupervised learning of semantic relations between concepts of a molecular biology ontology. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 2005. accepted for publication.

[9] I. Dagan, O. Glickman, and B. Magnini, editors. *The PASCAL Recognising Textual Entailment Challenge.*, 2005.

[10] I. Dagan, O. Glickman, and B. Magnini. The pascal recognizing textual entailment challenge. In *Proceedings of the PASCAL workshop on Recognizing Textual Entailment*, Southampton, UK, April 11-13, 2005.

[11] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D.S. Weld, and A. Yates. Web-scale information extraction in KnowItAll (preliminary results). In *Proceedings of the 13th World Wide Web Conference*, pages 100–109, 2004.

[12] R. Evans. A framework for named entity recognition in the open domain. In *Proceedings of the Recent Advances in Natural Language Processing (RANLP-2003)*, pages 137–144, 2003.

[13] D. Faure and C. Nedellec. A corpus-based conceptual clustering method for verb frames and ontology. In P. Velardi, editor, *Proceedings of the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications*, pages 5–12, 1998.

[14] K. Frantzi and S. Ananiadou. The c-value / nc-value domain independent method for multiword term extraction. *Journal of Natural Language Processing*, 6(3):145–179, 1999.

[15] P. Gamallo, M. Gonzalez, A. Agustini, G. Lopes, and V. S. de Lima. Mapping syntactic dependencies onto semantic relations. In *Proceedings of the ECAI Workshop on Machine Learning and Natural Language Processing for Ontology Engineering*, 2002.

[16] A. Gomez-Perez and D. Manzano-Macho. A survey of ontology learning methods and techniques. deliverable 1.5, ontoweb project, 2003.

[17] G. Grefenstette. *Explorations in Authomatic Thesaurus Construction.* Kluwer, 1994.

[18] T. Gruber. Towards principles for the design of ontologies used for knowledge sharing. *Int. J. of Human and Computer Studies*, 43:907–928, 1994.

[19] Z. Harris. *Mathematical Structures of Language*. John Wiley & Sons, 1968.

[20] M.A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 539–545, 1992.

[21] D. Hindle. Noun classification from predicate-argument structures. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 268–275, 1990.

[22] J.-U. Kietz, A. Maedche, and R. Volz. A method for semi-automatic ontology acquisition from a corporate intranet. In *EKAW'2000 Workshop on Ontologies and Text*, 2000.

[23] T.K. Landauer and S.T. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104:211–240, 1997.

[24] M. Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *ACM SIGDOC '86, The Fifth International Conference on Systems Documentation*, 1986.

[25] D. Lin and P. Pantel. Dirt - discovery of inference rules from text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 323–328, 2001.

[26] D. Lin and P. Pantel. Induction of semantic classes from natural language text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 317–322, 2001.

[27] D. Lin and P. Pantel. Concept discovery from text. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 577–583, 2002.

[28] A. Maedche. *Ontology Learning for the Semantic Web*. Kluwer Academic Publishers, 2002.
[29] Alexander Maedche and Steffen Staab. Discovering conceptual relations from text. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intellignece (ECAI'2000)*, 2000.
[30] Bernardo Magnini and Carlo Strapparava. Experiments in word domain disambiguation for parallel texts. In *ACL workshop on Word Senses and Multilinguality*, 2000.
[31] R. Navigli and P. Velardi. Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics*, 50, 2004.
[32] P. Pantel and D. Lin. A statistical corpus-based term extractor. In E. Stroulia and S. Matwin, editors, *AI 2001*, Lecture Notes in Artificial Intellingence, pages 36–46. Springer Verlag, 2001.
[33] J. Pustejovsky, J. Castano, J. Zhang, B. Cochran, and M. Kotecki. Robust relational parsing over biomedical literature: Extracting inhibit relations. In *Pacific Symposium on Biocomputing*, 2002.
[34] P. Resnik. Selection and information: A class-based approach to lexical relationships, 1993.
[35] T. Rindflesch, L. Tanabe, J. Weinstein, and L. Hunter. Edgar: Extraction of drugs, genes, and relations from biomedical literature. In *Pacific Symposium on Biocomputing*, 2000.
[36] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):515–523, 1988.
[37] M. Sanderson and B. Croft. Deriving concept hierarchies from text. In *Research and Development in Information Retrieval*, pages 206–213. 1999.
[38] Hinrich Schütze. Word space. In *Advances in Neural Information Processing Systems 5*, 1993.
[39] D. Turcato, F. Popowich, J. Toole, D. Fass, D. Nicholson, and G. Tisher. Adapting a synonym database to specific domains. In *ACL'2000 Workshop on Recent Advances in Natural Language Processing and Information Retrieval*, 2000.
[40] Peter D. Turney. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proceedings of the 12th European Conference on Machine Learning*, pages 491 – 502, 2001.
[41] Spela Vintar, Ljupco Todorovski, Daniel Sonntag, and Paul Buitelaar. Evaluating context features for medical relation mining. In *ECML/PKDD Workshop on Data Mining and Text Mining for Bioinformatics*, 2003.
[42] D. Yarowsky. Word-sense disambiguation using statistical models of roget's categories trained on large corpora. In *COLING-92, Nantes*, 1992.